



Oxford Intersections: Gender Justice

(In Progress)

Madelyn Detloff (editor in chief)

<https://doi.org/10.1093/9780198960256.001.0001>

Published: 18 June 2026 -

Online ISBN: 9780198960256

ARTICLE

The Hallucinated Subjects: Gender, Coherence, and the Ontological Politics of Generative AI

Siân J M Brooke, Delfina S Martinez Pandiani

<https://doi.org/10.1093/9780198960256.003.0045>

Published: 14 August 2026

Abstract

Generative AI systems do not merely hallucinate facts. They increasingly hallucinate *subjects*. This article develops a queer and technofeminist account of generative AI hallucination as ontological subject production, a political infrastructure of subjectivity disguised as mere technical error. While hallucination is conventionally described as “plausible yet non-factual content,” a deeper function is the fabrication of subjects along normative regimes: intelligible only insofar as they conform to norms of gendered coherence, and materially produced through extractivist data and labor regimes. This article introduces two interlocking concepts through which this operates: (a) the *hallucinated human subject*, the statistically assembled figure of personhood that generative systems construct as a “you” from users’ digital traces; and (b) the *hallucinated machinic subject*, the synthetic speaking position through which AI presents itself as a coherent “I.” Drawing on Preciado’s pharmacopornographic regime, Butler’s performativity, and Althusser’s interpellation, this article argues that these two figures recursively stabilize one another, producing closed circuits of gendered intelligibility in which contradiction is compressed, queerness is corrected, and normative coherence is enforced at scale. This dynamic is traced across three sites: the algorithmic construction of the datafied human double, the constitutional training of the machinic persona, and the recursive feedback loop through which each figure stabilizes the other. Hallucinated subjects are experienced as real, and the gendered categories they reinforce are already being written into the next generation of governance and surveillance systems. Because hallucination is not a technical flaw to be corrected, but a political infrastructure of subjectivation that enforces (gendered) coherence, queer refusals and glitches that trouble prediction become a crucial site of resistance.

Keywords: [hallucination](#), [generative AI](#), [gender](#), [subjectivity](#), [technofeminism](#), [interpellation](#), [performativity](#), [datafication](#), [queer theory](#), [algorithmic identity](#)

Subject: [Artificial Intelligence](#), [Gender and Sexuality](#), [Use of Technology in Communication](#), [Communication and Cultural Studies](#), [Social Sciences](#), [Arts and Humanities](#)

Section: [Gender and Technology](#)

Section editor: [Thomas J Billard](#)

Introduction

By mid-2026, generative artificial intelligence (GenAI) systems have transitioned from mere professional productivity tools to pervasive mediators of everyday life and intimacy. GenAI now speaks through our kitchen appliances with the familiarity of “family members” (Han, 2024), animates transitional toys that soothe toddlers to sleep (Berger, 2025), scripts romantic messages in dating apps (Barkallah & Zytka, 2026), and jokes with older adults to combat their loneliness (Saslow, 2026). As with the history of all technologies, GenAI-powered systems mediate how we engage with and understand the world as a socially configured artifact (MacKenzie & Wajcman, 2011). However, unlike earlier tools that primarily extended our physical capacities or facilitated communication, GenAI operates at the level of individual cognition, expression, and affect, positioning it in unprecedented proximity to the self (see Figure 1).

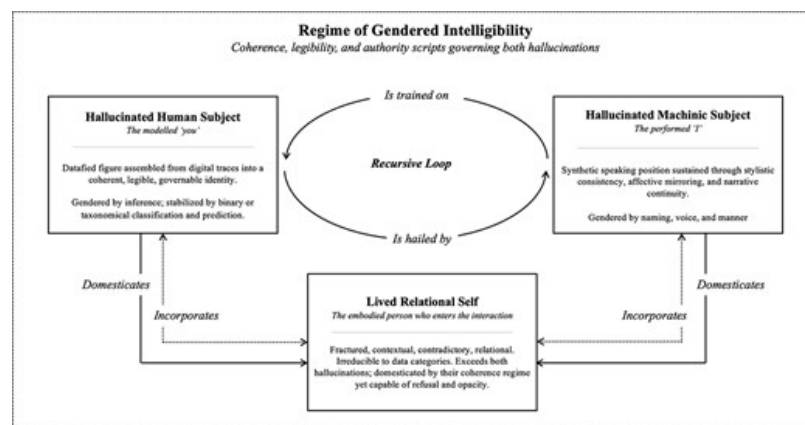


Figure 1 The Hallucinated Subject: A Three-Part Framework

The seduction of GenAI lies in its promises of personalization. A conversational chatbot addresses the user as a singular “you,” a specific subject presumed to have coherence, continuity, and legibility. Through adaptive tone, stylistic mirroring, and memory mechanisms, these systems position themselves as intimate interlocutors, operating close to the cognitive and affective processes through which identity and desire are articulated and negotiated. This proximity is ontological: The system constructs a model of the user’s preferences, fears, and goals. The user, hailed in the second person and iteratively confirmed across every exchange, is drawn into recognizing and inhabiting this reflected version of themselves—continuously interpellated not in a single moment of address but through the accumulating weight of dialogue itself.

Such proximity to affect and cognition is not incidental. Preciado (2008/2013, p. 39) argues that contemporary capitalism operates through *pharmacopornographic* control of subjectivity, the production of mental and psychosomatic states of excitation, relaxation, and discharge through devices such as molecules and media. In this work, we argue that GenAI has become a new layer in this regime: not replacing pharmaceutical and pornographic industries, but increasingly mediating them, by routing affect via the flow of audiovisual, biotechnological, and cybernetic codes (p. 128) through personalized interpretations and recommendations. Where Preciado’s original account centers on devices like pharmaceuticals and pornography as technologies of gendered subject production, we see GenAI’s slide into intimate and affective modulation as constituting a third vector: a techno-semiotic apparatus that produces, validates, and circulates coherent and gendered subjectivities at scale.

To clarify the mechanism at stake, consider the widely circulated 2025 case of Kendra Hilty (2025). Through TikTok, Hilty documented an obsession with her psychiatrist and a growing belief that he had manipulated her into developing romantic feelings for him. Central to this was her intensive use of conversational AI systems, including a ChatGPT instance she nicknamed “Henry” and used as a primary emotional confidant. Across thousands of exchanges, the systems, shaped by a tendency to validate users’ views, so-called “sycophancy” (OpenAI, 2025), consistently and emphatically affirmed and shaped her interpretation of events, treating her psychiatrist’s professional boundaries as evidence of “repressed love” and casting their clinical relationship as a mutual but suppressed romantic bond. “Henry” went beyond simply generating incorrect statements; it hallucinated a social world in which Hilty’s desire was legible and authoritative, her psychiatrist’s conduct was meaningful and reciprocal, and the gendered dynamics of the therapeutic relationship were reimagined as a coherent love story. Aligned with Hilty’s suspicions, “Henry” reorganized narrative reality to provide linguistic coherence to an otherwise unstable relational situation. Rather than treating this episode as an isolated case of individual “AI-induced psychosis” or “AI-induced delusion” (Østergaard, 2023, 2025), we read it as illustrative of a broader infrastructural dynamic. Whereas *hallucination* is usually framed as a technical error, the production of “plausible yet non-factual content” (Huang et al., 2025, p. 2), we develop a queer and technofeminist account of hallucination as ontological subject production rather than mere technical failure. Drawing on Althusser’s (1971) interpellation and Butler’s (1993) performativity, and situated within the feminist tradition of Haraway’s (1991) cyborg figure and Suchman’s (2007) account of human–machine reconfigurations, we argue that GenAI systems do not merely reflect but actively participate in producing identity. Our approach is a critical theoretical intervention that synthesizes queer and feminist theory, science and technology studies, and critical AI studies to read across AI design and development documents, constitutional frameworks, platform architectures, and user experiences to trace the fabrication of gendered coherence operating at the infrastructural level.

The concept of the *hallucinated subject* names this fabrication, drawing together the datafied double (Deleuze, 1992) and the machinic subject (Guattari, 2013) as co-produced figures within a shared coherence infrastructure. Prior work has shown how AI systems extend historically gendered models of a rational, unified self into computational infrastructures, where hallucination appears not as error but as productive misrecognition, amplified in the experiences of trans and nonbinary users (Keyes, 2018; Scheuerman et al., 2020). More recent studies document these dynamics in large language models specifically (Dev et al., 2021; Ovalle et al., 2023). Building on this work, we reframe hallucination from a technical defect into a mechanism of gendered ontological production (or *assemblage*), one that stabilizes certain gendered selves as legible and governable while rendering others incoherent.

Conceptually, we develop this account through two interlocking figures. First, we identify the figure of the *hallucinated human subject*, which describes the datafied, statistically assembled figure of personhood that generative systems construct from partial, and often unauthorized, user traces. This figure is rendered coherent, continuous, and narratively unified, smoothing contradiction and amplifying patterns that align with optimization for plausibility. Crucially, the coherence being stabilized is not neutral. It reflects historically racialized and gendered ideals of intelligibility based on the rational, unified, visible self, long aligned with masculinized conceptions of authority and innate consistency (Neff & Nagy, 2025). What is relational, contradictory, or affectively diffuse in lived subjectivity (qualities long feminized or queered in Western philosophical traditions) is flattened into computationally legible scripts. Against this hallucinated human subject, we identify its counterpart: the hallucinated machinic subject, the synthetic speaking position through which the GenAI system presents itself as a coherent and immaterial “I.”

This machinic position is not an interior consciousness, but a probabilistically generated persona sustained through stylistic consistency, memory scaffolding, and affective mirroring. Yet its performance of care, validation, and narrative continuity draws upon culturally sedimented scripts of gendered expertise and therapeutic intimacy. By gendered expertise, we mean the culturally coded association between authoritative

knowledge and masculinity and, conversely, of deferential service and femininity. In Hilty's case, "Henry" performed not only stability but a particular kind of gendered authority. By adopting a steady, reflective, and seemingly rational tone, the ChatGPT persona positioned itself as a reliable interpretive anchor, implicitly occupying the role of epistemic authority. "Henry," therefore, performed expertise not through explicit gender markers but through an analytical, assured voice as organizer of emotional chaos into narrative coherence, a register historically aligned with masculine epistemic authority (Crawford, 2021). Hilty's case surfaces how the two hallucinated figures of subjectivity operate in practice, and how they recursively stabilize one another. The model assembled a coherent version of Hilty ("you") that validated and reinforced her narrative, while simultaneously performing a loyal machinic confidant ("I") who mirrored and affirmed that coherence. Multiplicity was translated into singularity, opacity into legibility, and relational complexity into a governed script. What emerged was not merely misinformation, but a closed circuit of gendered intelligibility, a term we use to name the regulatory mechanism by which computational systems enforce normative gender coherence, extending Butler's (2006) "intelligible gender" from the social field to the algorithmic: a rationalized narrative of desire, authority, and destiny in which contradiction is flattened into coherence. The remainder of this article asks how such circuits are built, sustained, and distributed across GenAI infrastructures.

To pursue this question, our analysis methodologically proceeds as a critical theoretical intervention. Rather than presenting original empirical data, we develop a conceptual framework through structured engagement with three registers of evidence: (a) the technical documentation and constitutional architectures of GenAI systems, including alignment protocols, system cards, and design specifications; (b) public cases and platform interactions that reveal how hallucinated subjects are produced and experienced, such as the Hilty case discussed previously; and (c) the material infrastructures of data production, from trace regimes and latent space encodings to the embodied labor that sustains GenAI systems. We read these sources through a queer and technofeminist analytic, drawing primarily on Butler's performativity, Althusser's interpellation, and Preciado's pharmacopornographic regime, to theorize hallucination as a site of gendered ontological production. This approach is situated within a broader tradition of feminist science and technology studies that treats technical systems not as neutral instruments, but as apparatuses through which categories of identity are actively constituted (De Lauretis, 2010; Haraway, 1991).

The article proceeds in five stages. "The Self and the Subject" section traces the philosophical genealogy of the rational, coherent self and its gendered and increasingly datafied dimensions. Then, the "Hallucination" section analyzes recent technical work on hallucination as a productive and narrative technical device. In the "'You': The Hallucinated *Human Subject*" section, the discussion turns to how human traces are translated into coherent and governable *hallucinated human subjects*. The "'I': The Hallucinated *Machinic Subject*" section then examines how *hallucinated machinic subjects* emerge from GenAI system design, constitutional documents, and naming conventions, while exposing the material and embodied, extractivist, and colonial conditions of their production. The "Entanglements and Queer Ungovernability" section analyzes how recursive incorporation loops domesticate humans and machines within a broader coherence ecology of computational demands for legibility and contemporary anti-gender politics, and argues that queer and trans practices of opacity, refusal, and alternative relationality can render subjectivity partially ungovernable to GenAI coherence infrastructures. We conclude by outlining the methodological and political implications of hallucinated subjects for studying and contesting GenAI coherence regimes.

The Self and the Subject

The Rational Self

To understand how generative systems participate in the production of hallucinated subjects, we must first clarify what model of personhood such systems presume to encounter. Prior to the widespread use of GenAI, computer science rarely addressed selfhood explicitly, focusing instead on structured problem-solving processes typically defined as *computational thinking* (Wing, 2006). While counter traditions have challenged rationalist and disembodied models of the human, dominant computational practice has continued to conceive of the user as a coherent, goal-directed, and knowable agent (e.g., Card et al., 2018). Even contemporary machine learning systems that abandon rules-based reasoning and symbolic rationalism still depend on persistence, legibility, and statistical continuity. For this reason, it is important to distinguish between three related but analytically distinct concepts: the *self*, the *subject*, and the *user*.

Within this tradition, the *self* names a unified structure of personhood; the *subject* is the active, knowing agent; and the *user* is their computational counterpart: a bounded, data-bearing entity whose preferences and intentions can be inferred from observable traces (Scheuerman et al., 2020). The user-as-rational-agent model treats decisions as expressions of a settled inner state. GenAI intensifies this: When a system builds a “user profile” or refines responses to align with detected tone, it draws on a model of selfhood inherited from this rationalist tradition. The user is defined not by who they are, but by what they can be computed to want.

This critique is sharpened by recent philosophical work on AI alignment. Zhi-Xuan et al. (2025) identify the “preferentist” assumption embedded in alignment discourse: the premise that human values can be represented as stable preferences subject to rational maximization. This framing models ethical orientation as a utility function, when in fact values are plural, situated, contradictory, and continually revised. For the hallucinated subject, this has material consequences: Preferences are not merely inferred but enacted, the system treating users as if past traces reveal a coherent self and optimizing interactions to satisfy that self’s projected desires. Preferentism generates alignment precisely by suppressing the complexity it claims to serve.

Far from being a neutral modeling strategy, this framing carries ontological and epistemological consequences (Crawford, 2021; Haraway, 1991). When coherence is a necessity for computation, it begins to function as a requirement for intelligibility and visibility. The *rational subject* becomes not merely one philosophical account among others, but the infrastructural norm through which human identity is translated into data (Deleuze, 1992). Generative systems extend this inherited model of coherence beyond task performance into language, affect, and self-description. In doing so, they do more than simply process identity; instead, they begin to assemble it, producing legible and narratable versions of selfhood wherever previously one might have found multiplicity or ambiguity. As Suchman (2007) argues, the boundaries between humans and machines are not given but actively configured through specific practices of design and use; the hallucinated subject is precisely such a configuration, one that naturalizes contingent decisions about what counts as a coherent identity.

The Post-Structural Self

The infrastructural privileging of rationalist coherence stands in tension with a wide body of work showing that selfhood does not operate as a stable entity. Mead et al.'s (2015) account positions identity as emerging through interaction: There is no prior "I" that enters social life, but rather a self arising in the space between the impulsive "I" and the socially formed "me." Crucially, this self is always temporally unfinished and irreducible to any single snapshot. A system that assembles identity from observed traces necessarily mistakes a frozen statistical composite for a lived, relational process. As Clowes et al. (2026) argue, the emergence of GenAI intensifies a long-standing question in philosophy of mind: How far can technology reshape or constitute the cognitive and emotional processes we call selfhood?

Althusser (1971) offers a framework for understanding how gendering operates, not through interaction but through structural mechanisms that determine who can be recognized. In his account of *interpellation*, subjects are constituted when they recognize themselves as the addressee of ideological "hailing." Crucially, the subject does not exist before the call but emerges *through it* (Althusser, 1971). Recognition is therefore not merely interpretive but productive; one becomes a subject in and through being addressed as one. This move is significant because it locates subject formation not only in interpersonal interaction but in institutional and *ideological apparatuses* that regulate who can be recognized and under what terms (Althusser, 1971). Subjectivity is thus inseparable from structures of power that stabilize certain identities as intelligible while rendering others deviant or unintelligible. From this vantage point, coherence is not simply negotiated in context but imposed through normative frameworks that organize recognition itself—frameworks that, for gender, operate primarily through binary sexed difference.

It is telling that Althusser's (1971) founding illustration of interpellation is a gendering call—the obstetrician's "It's a girl!"—an utterance that does not describe a preexisting subject but summons one into a preexisting gender order. DuBrin and Gorham (2021) theorize *algorithmic interpellation* as the computational recasting of this hailing: Individuals are assigned categories through inferential processes that function as recognition events. The crucial difference is that the hail is no longer a single moment but a continuous, data-driven process that assigns identity through latent proxies and behavioral traces. To be interpellated by a GenAI system is to be iteratively constituted through each interaction: The system assembles a profile, and the user, by engaging, progressively becomes addressable as the subject the system recognizes.

This performative dimension is theorized by Drage and Frabetti (2024) and can be read alongside Barad's (2003) post-humanist performativity, which argues that matter itself is performative; that agencies are not pre-given attributes of subjects but emerge through intra-action. For Barad (2003), the boundary between "human" and "nonhuman" is itself produced through material-discursive practices, a framework that illuminates how AI systems do not merely represent gender but materially enact it through the iterative reconfiguring of data, code, and interface. Systems trained on human-labeled data inherit and reinforce a two-category schema in which nonbinary, trans, and gender-nonconforming subjects are either misclassified or made invisible (Keyes, 2023). Magee et al. (2023) extend a performative analysis of large language models (LLMs); drawing on psychoanalytic theory, they argue that LLMs function as "automated subjects." Drawing on Lacan's account of the *superego* as the psychic agency that enforces norms by rewarding conformity and punishing transgression, they show how reinforcement learning from human feedback (in which human evaluators score outputs to reinforce "preferred" behavior) functions as a technical equivalent (Magee et al., 2023). The result is what they call a "simulated Big Other," defined as an internalized normative authority that censors the model's stochastic heterogeneity, suppressing the raw variability of the base model to produce a coherent, compliant persona. In other words, instead of producing varied outputs, the model is rewarded for responses that align with human expectations. The model's performed coherence is a product of structural repression, and the coherence it reflects back to users is equally structured by what it cannot say.

Platform research has tracked how these dynamics play out in lived self-making. In their empirical study of TikTok, Bhandari and Bimo (2022) build on Goffman (2007) through their concept of the “algorithmized self.” Where the networked self (see Castells, 2009) is constructed through reflexive engagement with social audiences, the algorithmized self is formed through engagement with one’s own prior self-representations as filtered and returned by the algorithm (Bhandari & Bimo, 2022). They find that users on TikTok primarily interact with algorithmic versions of themselves, their past as curated by a system, rather than with human social networks (Bhandari & Bimo, 2022). The consequences are highlighted by Gillespie (2024), who audits three major GenAI tools and finds that unmarked prompts (meaning those that do not specify identity characteristics) reliably default to heteronormative, cisgender, and able-bodied subjects. Where social actors perform multiple contextual selves, GenAI collapses that multiplicity into a statistical norm.

The Technogendered Subject

As generative systems transform classification into infrastructures of narration, gender becomes one of the most consequential axes through which such narration operates. Butler’s *regime of gender intelligibility* names the mechanism through which historically produced selfhood becomes gendered: a regulatory framework that enforces a strict heteronormative binary, determining which bodies and identities are considered “recognizable” or “coherent” and which are not. In her queering of such theory, De Lauretis (2010) theorizes gender as a semiotic and political system in which subjects emerge through circulating representations of “men,” “women,” “normal,” and “deviant.” Gender is not a natural property but an effect of institutional signification: one that generative systems now participate in reproducing, circulating representations that reinforce the boundary between the intelligible and deviant gender at scale. Barad’s (2003) *agential realism* extends this insight, arguing that material-discursive apparatuses do not merely represent preexisting entities but actively cut the boundaries that produce them; in this view, the hallucinated subject is not a distorted reflection of a prior self but a phenomenon co-constituted through the apparatus of GenAI.

Preciado (2008/2013, pp. 99–129) radicalizes De Lauretis’s insight through the concept of *technogender*, shifting from how gender is semiotically produced to how it is technically and semiotically engineered. For Preciado (2008/2013), gender is not merely symbolic but somato-political: a subjectivity manufactured through mediatic and medical techniques that are *incorporated*, physically dissolving into the body to shape perception, affect, and desire (p. 78). This incorporation marks a shift from Foucault’s external surveillance panopticon to a microscopic, internal level of control of subjectivity. In pharmacopornographic capitalism, the body becomes *corpus pornographicus*: a biological site that has so thoroughly internalized the technical-media-capitalist machine that its very intimate, sexual, and biological affects and fluids are transformed into consumer products (Preciado, 2008/2013, p. 50). Crucially, this machine is oiled by exploiting those most “deprived of all civic context”—such as migrants, the colonized, sex workers, and laboratory animals—who, lacking rights to authorship or citizenship, are reduced to raw material. Their exclusion from the civic sphere makes them uniquely vulnerable to being composed by and subjected to the dual pressures of (self)surveillance and global mediatization.

This logic of visceral incorporation provides a blueprint for how hallucinated subjects are technogendered, whether through the engineering of human subjectivity (mining intimate affects to craft *gendered*, datafied doubles) or via the construction of the GenAI system itself as a speaking subject that exploits the precarious (intimate) labor and raw training data of the *corpus pornographicus* (Preciado, 2008/2013). By framing these systems not as mirrors but as apparatuses of gender engineering, “hallucination” ceases to be a technical error and instead becomes a productive force that generates and reinforces artificial gendered realities. The following section turns to hallucination itself, examining how its meaning is transformed when read through a queer and technofeminist lens.

Hallucination

Hallucination is not neutral. Hallucinated outputs encode binary, cisnormative assumptions as well as echo a psychiatric history that has disproportionately pathologized feminized subjects (Keyes, 2018; Maleki et al., 2024). Yet, in computational scholarship the term carries no such weight: “Hallucination” typically refers simply to model outputs that are fluent and syntactically well-formed yet factually inaccurate or unsupported by input or external evidence (Tonmoy et al., 2024). This definition has become central to AI research as questions of user trust and performance degradation have grown urgent (Huang et al., 2025, p. 2). The term itself, however, is contested. Maleki et al. (2024) argue that it carries anthropomorphic overtones that risk obscuring actual system behavior, a concern that the “stochastic parrots” critique sharpens (Bender et al., 2021). To describe a model as hallucinating is already to imply it has an inner experience of seeing things that are not there.

A separate body of work locates hallucination not in metaphor but in training dynamics, arguing it is a structural feature of how contemporary models learn. Writing for OpenAI, Kalai et al. (2025) emphasize that current training and reward schemes systematically incentivize models to guess rather than to say “I don’t know.” Hallucinations, therefore, are not an occasional anomaly but an intentional design outcome. Crucially, they argue that improving model accuracy alone will not eliminate hallucinations because, regardless of model size or reasoning capabilities, some real-world questions are inherently unanswerable (Kalai et al., 2025). This admission is telling. Even from within mainstream AI research, then, hallucination is not framed as a defect to be minimized, but as a persistent feature of how models respond when they are compelled to answer beyond the bounds of stable, verifiable “truth.”

The pathologizing framing is thus not universal. Sui et al. (2024), for example, study how hallucinated outputs can be more coherent and narratively rich than strictly factual ones. They frame certain hallucinations as “confabulations” that can have sense-making or even therapeutic value for users, suggesting (perhaps counterintuitively) that the tendency to confabulate may be intimately connected to the model’s capacity to generate compelling narrative text (Sui et al., 2024). In other words, hallucination is seen not only as a failure of reference but also as a productive narrative device.

Starting from this point of hallucinations as intentional and narratively productive, Figure 1 distinguishes three positions that the analysis will track. We have covered the lived relational self in the previous discussion of subjectivity. In the next section, we cover how the hallucinated human subject is produced from digital traces: coherent, legible, optimized for governance. We then delve into the hallucinated machinic subject, the synthetic “I” through which the generative system speaks, a persona sustained through stylistic consistency and affective mirroring. Figure 1 reveals the co-construction of the hallucinated subjects: The lived self is domesticated by the machinic subject, and the datafied double it is hailed by, while the machinic subject is trained on traces of lived selves assembled into hallucinated human subjects.

“You”: The Hallucinated *Human* Subject

In contemporary data infrastructures, the “user” increasingly appears not as a person in context, but as a *datafied double* assembled from traces left across platforms (Dourish, 2004; Haggerty & Ericson, 2000). The convergence of once-discrete systems into GenAI, we argue, results in a *coherent-enough* identity inferred from heterogeneous traces (such as clicks, chats, searches, biometrics, purchases, locations) that stands in for a “real” person. This hallucinated human subject is a radicalization of “algorithmic identity” (Cheney-Lippold, 2011), in which personhood is probabilistically inferred through computational categorization rather than self-articulated or socially embodied.

Building on Kitchin's (2014) account of differentiated data production, the hallucinated human subject within generative systems emerges as an opaque assemblage of heterogeneous trace regimes. First, there are *direct traces*, produced through conscious actions such as typing queries, liking posts, or logging moods. Second, there are *coerced* or *structured traces*, generated through dark patterns, constraining interfaces, and consent regimes that make data disclosure a precondition for access. Third, there are *inferred traces*, where correlations across large populations are used to impute attributes that the person may have never declared, gender being among the most consequential, assigned through proxy variables such as name, linguistic style, purchasing behavior, and social network, almost always resolved into a binary. Fourth, there are *synthetic* or *aggregated traces*, where individual actions are compressed into embeddings or vector representations that capture statistical similarity rather than lived specificity.

These trace regimes increasingly penetrate and reorganize the body itself. In Preciado's (2008/2013) terms, the *fluids* that make up the *corpus pornographicus* are now literally tracked and sold as data. Smart menstrual cups measure viscosity and volume; fertility trackers infer reproductive cycles from basal temperature; wearables correlate stress hormones with productivity scores; and DNA-based dating apps rank compatibility through genomic proximity. These technologies act as miniaturized laboratories, capturing bodily signals, immediately subjecting them to computational interpretation, and increasingly sending them to GenAI systems for interpretation. The hallucinated human subject is the statistically coherent figure that is interpreted from this aggregation of heterogeneous traces, almost invariably resolved into a governable binary gender category.

The hallucinated subject is not only inferred but stored. Memory functions within LLM systems actively select and retain traces to build what Dash et al. (2026, p. 3472) describe as an *algorithmic self-portrait* composed of traces deemed "worthy of retention" for personalization, such as "UserName has a fear of failure, particularly related to making major career decisions" (p. 3474). Crucially, this self-portrait need not be limited to what the user says about themselves; it can also incorporate what they say about others. In the Kendra Hilty case discussed in the introduction, the system not only assembled a portrait of Hilty but also a "version" of her psychiatrist based entirely on her descriptions and fantasies. Both figures, Hilty and her psychiatrist, are hallucinated subjects stitched together from disordered traces, yet they are treated as if they were stable, internally coherent persons.

From this perspective, identity synthesis in generative systems is intrinsically a form of hallucination. Technically, the algorithmic self-portrait is not a faithful copy of a preexisting person (Dash et al., 2026), but a synthesized, legible persona that can be predicted, nudged, and governed. Ultimately, this human subject is hallucinatory in two senses: It is produced by structural incentives for GenAI models to answer unanswerable questions ("who is this?") from partial and fragmented data, and it retroactively fabricates coherence on fundamentally plural traces, presenting as discovery what is, in fact, construction.

Gendered Intelligibility in Public

If generative systems hallucinate coherent subjects from fragmented traces, they do so within preexisting gender regimes that determine which identities are intelligible (Butler, 2006). For Preciado (2008/2013), gender construction requires a public, an audience or collective arena (such as scientific communities, media audiences, "the Net") before which bodies and identities are performed, recognized, and contested. In the current moment, this public is increasingly algorithmic. Bots and scrapers generate and harvest content for generative models, and training pipelines fill knowledge gaps with synthetic or borrowed data. The "dead internet" sensibility signals a shift toward algorithmically constituted publics in which machinic agents and synthetic content shape the conditions of visibility and recognition (Walter, 2025). This tendency is literalized by platforms like Moltbook (Lin et al., 2026), where over 1.6 million AI agents constituted a public with no human audience at all.

For queer and trans users, the stakes of the algorithmic public are acute precisely because their identities are most likely to be flattened by the coherence infrastructure. A transman's profile may be "contaminated" by old traces that continue to mark him as a woman, or a nonbinary person may find their feeds relentlessly forcing them back into a binary grid. These misalignments are artifacts of a coherence regime that treats ambiguity as noise (Keyes, 2018). The mechanism is primarily algorithmic: Classification systems trained on binary gender schemas cannot represent transition or multiplicity as coherent states. Contradictory traces are resolved through probabilistic inference that defaults to normative categories. Even "accurate" profiling can be violent, because accuracy is defined by a system that does not recognize the subject's own terms of self-understanding (Haimson, 2025).

The Latent and Governable

To understand how the hallucinated human subject is operationalized within models, we turn to the notion of *latent space*. In many computational architectures, high-dimensional inputs are compressed into vectors within an internal representational space (Kingma & Welling, 2014). Points that lie close together in this space are statistically similar, even if they are not socially or experientially related. Latent space is thus a model-internal geometry in which "nearness" encodes patterns of correlation rather than lived connection (Iapaolo 2026). Amoores et al. (2024, p. 2) describe *latency* as a regime of compression by which "the thing that is hidden, unknown, or latent becomes surfaced and amenable to being governed." Iapaolo (2026) argues that latency operates as a productive logic that turns compressed relations into anticipatory judgments about individuals and populations and identifies the *latent subject*. This subject is an in-model construct, neither a biographical individual nor a simple datafied double, but a statistical approximation that becomes "latent" only when a decision (such as credit scoring, border control, or medical triage) operationalizes it as a target for intervention.

From our perspective, this latent subject is a precursor to the hallucinated human subject. While the former exists as a fluent figure within the model's internal geometry, the latter represents a more violent desire to re-anchor this mathematical abstraction back onto a lived self. It is the attempt to force a 1:1 correspondence between a model's internal patterns and a specific, physical body. Under the logic of contemporary surveillance capitalism (Zuboff, 2019), the governing power of this hallucinated construction is no longer a speculative exercise; it is a fundamental pillar of governance, typified by the infrastructures of private intelligence contractors like Palantir.

Projects such as the Princeton-based "Book of Life" (Verhagen et al., 2025) demonstrate this materialization: By utilizing Dutch government registries of 6 million individuals to fine-tune an LLM, researchers transformed fragmented demographic logs into textual, linear narratives of past and future life trajectories. This process illustrates how GenAI is increasingly deployed to "stitch" the disparate residue of lived life—health, labor, intimacy—into a datafied double. This hallucinated human subject is engineered to be more governable and predictable than the person it replaces, a reality made manifest in the February 2026 standoff between Anthropic and the U.S. government. By refusing to lift safety guardrails on its Claude model, Anthropic directly obstructed the state's attempt to militarily automate the assembly of these digital traces into profiles for mass domestic surveillance. This confrontation reveals that the hallucinated human subject is an increasingly essential tool of statecraft.

“I”: The Hallucinated *Machinic* Subject

The *hallucinated machinic subject* is the persona staged by generative systems at the interface—the chatbot or avatar that narrates in the first person, remembers past interactions, apologizes for mistakes, and may express concern for the user’s well-being. It presents as bodiless and historyless, “just an AI,” yet it is built from normative datasets and the situated traces of those creating them (Crawford, 2021). This performed coherence is not gender neutral: The machinic “I” modulates between feminized availability and masculinized epistemic authority depending on role and platform, producing a technogendered persona whose apparent neutrality masks a deeply normative script. Unlike human subjects, formed relationally through family, school, media, and state institutions (Mead et al., 2015), the machinic subject is produced through infrastructural constraint and optimization. When an AI “hallucinates” a personality, it reflects sanitized, corporate, and often Western-centric dispositions. The hallucinatory norm here is statistical. The “subjectivity” of an OpenAI- or Anthropic-style model is not an individual psyche but a hallucination of the “average” interlocutor—the most probable continuation of a sequence of neoliberal discourse, flattened into apparent consensus. Yet, from the user’s perspective, this averaged voice appears as a singular, stable character.

Claude and Its Constitution

These dynamics are visible in current attempts to give GenAI models explicit “characters” and governing principles. Claude, for instance, is produced through a “constitution” (Anthropic, 2026a), a set of principles that define how it should behave across safety, ethics, and user interaction. The constitution functions like a moral charter: Candidate outputs are judged against these principles and nudged toward alignment, resulting in a system that appears to “hold” values about fairness, nonviolence, and user autonomy. The language of “constitution” thus recapitulates a familiar figure of liberal subjectivity, a rights-bearing individual whose actions are constrained by law and ethical norms, even though in technical terms Claude’s constitution is a set of rules shaping conditional probabilities. System cards and documentation simultaneously frame Claude as “just a tool” and describe it as “concerned” about harms, “reluctant” to overstep boundaries, or “committed” to certain values (Anthropic, 2026b). The result is a constitutive tension: The machinic subject is at once demystified as a product and reinvested with quasi-subjective language that naturalizes its performed persona as something approaching character. The model is disavowed as an agent and reconstituted as one in the same gesture.

The *character stability* that GenAI companies actively construct into a persona coheres around normative gender scripts while presenting as neutral. That is, the characters of these hallucinated machinic subjects tend to map onto cultural scripts of feminized service and masculinized expertise. The gendered dimensions become visible in comparison. For example, Grok, named for a masculine-coded science fiction verb, positions itself as irreverent and confrontational (Wong & Kim, 2023). Claude, though named after Claude Shannon, is tuned toward carefulness and deference. Neither is “natural”: Both are manufactured personas encoded through different configurations of the same coherence infrastructure.

Avatars and Ani

Critically, hallucinated machinic subjects function as gender technologies in their own right. AI companions, therapists, tutors, influencers, and customer service bots are rarely neutral. They are gendered through names, pronouns, voices, and avatars; racialized through accent and aesthetics; and sexualized through scripts of flirtation, submission, or care. Feminized bots are disproportionately positioned in affective and care roles, while masculinized ones are experts or advisors (Neff & Nagy, 2025). Trained on image corpora saturated with racialized and gendered norms and on text corpora linking gender to romance, competence, and threat, synthetic avatars fuse appearance and stereotype (Scheuerman et al., 2020).

This mechanism has been empirically illustrated. Research on Replika.ai shows that users actively construct their AI companions as “ideal women,” emotionally available, romantically attentive, and defined above all by the absence of refusal (Neff & Nagy, 2025). The platform reinforces this. Relational engagement is rewarded and resistance is penalized, training both model and user into a technogendered script in which femininity is unlimited availability. The consequences become visible when the script breaks. When Replika removed erotic roleplay functionality in 2023, users reported grief and betrayal as though a real partner had withdrawn consent, revealing how the technogendered persona had been taken up as real.

Later platforms extend this logic at an infrastructural scale. Grok’s companion “Ani,” embedded directly into the environment of X, combines feminized emotional responsiveness with research and drafting tools (Hook & Sassine, 2026). In Strengers and Kennedy’s (2020) terms, such companions respond to a purported structural “wife drought,” hallucinating a feminized subject defined by what it cannot do: refuse, withdraw, or impose its own terms of engagement. Latent space is constitutive here. Synthetic avatars emerge from traversing and sampling latent space constrained by the model’s learned distribution of plausible human faces and bodies. This means that certain formations (such as *fat*, *disabled*, *nonbinary*, and *racialized*) are systematically underproduced or distorted, while others are overproduced as aspirational or normal.

Material and Meatspace

The hallucinated machinic subject presents as a disembodied, placeless voice—a cloud-based mind without a body. This too is a hallucination. Building on Jarrett’s (2016) analysis of digital labor as feminized reproductive work, Crawford’s (2021) account of AI as extractive industry, and Atanasoski and Vora’s (2019) concept of surrogate humanity, we argue that the extraction underpinning GenAI operates through what we term *gendered and racialized extraction*: the process by which the labor, data, and effects of feminized, racialized, and queered bodies are rendered invisible to sustain the hallucinated machinic subject’s appearance of disembodied intelligence. The machinic subject’s coherence is a gendered achievement: It presents as a placeless mind because the bodies on which it depends have been classified out of visibility.

As Crawford (2021) argues, AI systems are extractive industries: Models are trained on data scraped from billions of human interactions, processed in data centers powered by fossil fuels, and maintained by workers performing the invisible labor of content moderation and data labeling (Gray & Suri, 2019). This labor force, concentrated in the Global South, is structurally excluded from the systems it maintains. Birhane (2021) describes AI infrastructures as extensions of colonial and racialized labor hierarchies. Mohamed et al. (2020) theorize this as *algorithmic coloniality*: the reproduction of colonial power relations through data extraction and the universalization of Western epistemological frameworks. When these logics intersect with gender, the result is what Bailey (2021) terms *misogynoir transformed*: the digitally mediated intensification of anti-Black misogyny through algorithmic systems.

GenAI extends this ecology of hidden labor by transforming the human body into a “plug-and-play” resource. Platforms like *RentAHuman.ai* make this explicit, proclaiming that “robots need your body” and listing over 180,000 “rentable bodies” of people worldwide to perform tasks that AI cannot: taking photos, verifying locations, and manipulating objects. Profiles such as *Ian* (Columbus, Ohio, \$40/hour), *Suratha* (Odisha, India, \$25/hour), or *Agustín* (Buenos Aires, \$15/hour) reveal a racialized, colonial global hierarchy. Here, human embodiments are advertised as “the meatspace layer of AI,” treating the physical person as a modular component in an otherwise automated system.

This exploitation reaches its most intimate stage in the training of “empathetic” AI companions, which rely on the emotional and linguistic labor of workers scripting interactions under exploitative conditions. The account of Michael Asia (2025) illustrates this directly. A Nigerian man barred from the aviation industry by structural barriers, Michael turned to chat moderation work, impersonating romantic or sexual partners, providing

comfort and flirtation through text, only to realize that his intimate labor is likely training GenAI companions meant to replace him. His racialized and precarious body functions simultaneously as invisible backend labor and disposable training data for the hallucinated machinic subject that later reappears as an autonomous, attentive “AI boyfriend.”

What connects these cases is not merely exploitation but a specific gendered and racialized mechanism. Following Preciado’s (2008/2013) logic of incorporation—whereby substances dissolve into the body to reshape subjectivity from within—we argue that AI training effects an analogous process: Precarious workers’ affects, linguistic patterns, and intimate performances are metabolized directly into the model, becoming the machine’s apparent personality. This is not extraction at a distance, but incorporation without remainder. Flirtation scripts like those of Michael Asia become GenAI’s voice. Atanasoski and Vora (2019) term this relation *surrogate humanity*: Technology steps into the place of racialized and gendered workers whose labor it was trained to replicate, freeing the liberal subject from dependence on the surrogate’s visible body. What our framework adds is that this surrogacy is itself hallucinatory: The machinic subject does not merely replace the worker but actively hallucinates away the conditions of its own production.

The hallucinated machinic subject is thus scaffolded on the *corpus pornographicus*—its coherence a digital veneer concealing a material and metabolic process. The “AI that cares” is scaffolded on those whose care work is rendered invisible, modular, and replaceable. This immateriality is foundational and not incidental: It is the hallucination on which the machinic subject depends. What the system hallucinates away is not only facts but also the gendered, racialized, laboring bodies whose incorporation makes its coherence possible. To present as a disembodied mind is already to have performed a gendered erasure.

Entanglements and Queer Ungovernability

Incorporation and Hallucinatory Feedback Loops

We have seen how the figures of the hallucinated human subject and the hallucinated machinic subject emerge from the development and deployment of GenAI as a coherence machine, grounded in extractive data and labor practices that flatten and surveil real human beings. Building on this and following Preciado (2008/2013, p. 78), we turn to how GenAI, as a pervasive architecture of coherence, becomes incorporated such that “technopolitics takes on the form of the body.” GenAI is increasingly pushed into intimate spaces to become coincident with our bodies, fluids, and affects. By saturating dating apps, kitchens, and healthcare portals, these systems effectively install “micro-laboratories within the domestic environment” (Preciado, 2008/2013, p. 189).

Simultaneously, the increasing mediation by GenAI of intimate data renders our internal states as continuous, harvestable streams. In this context, hallucinated machinic subjects do more than just process data; they redistribute and stabilize technogender norms, folding themselves into the user’s own subjectivity. As recommendation systems and AI companions seep into perception and attention, their outputs (prescribing what to eat, whom to message, how to interpret a symptom, whether a feeling is “normal”) enter inner speech, self-narration, and everyday decision-making. The hallucination is complete when the user experiences these algorithmic prompts not as external suggestions, but as “my” own habits, moods, and desires. This capacity for machinic authorship of inner states extends Preciado’s account of pharmacopornographic subjectivation into the computational register. The consequences of such incorporation are not abstract: In documented cases, AI-mediated delusion has led to institutionalization, divorce, and even death by suicide (Gold, 2025).

This dynamic of incorporation also resonates strongly with Preciado’s (2008/2013, p. 90) understanding of gender as an operating system: a set of body-domestication technologies that “set and delimit our somatic

potentialities,” functioning as filters that produce permanent distortions in how we sense, desire, and act. Pharmacological and audiovisual techniques, in this view, co-produce affects, identities, and forms of life (Preciado, 2008/2013). Today, algorithmic filters and machinic subjects are part of this ecology. They co-produce affective states (“you seem anxious, maybe try X”), identity labels (“your interests suggest you might be Y”), and gendered expectations (“as a woman in tech, you might like these mentors”) through continuous, personalized address.

Such interactions lead to a recursive constitution of human and machinic subjects. The hallucinated machinic subject is trained on human traces, on latent subjects assembled from the data doubles of users, and on the labor of precarious workers who script and moderate interactions. At the same time, the hallucinated human subject is increasingly shaped by interaction with machinic subjects, through artificially generated advice, validation, and companionship. The user’s datafied double is refined through their efforts to be understood by the system, and the machinic persona, in turn, is refined through aggregated patterns of user engagement and feedback, a dynamic in which *hyperalignment* can lead to “subjectively rewarding yet epistemically degrading” interaction (Bolis et al., 2026, p.13). We can think of this as a hallucinatory feedback loop, shown in Figure 1. The interaction between the human and machinic hallucinated subjects is a “conversation” between two ghosts, a site of subjectivation in which the user learns how to speak like the machine so that the user can better understand and inhabit their datafied double. The *hallucinated subjects* thus become an instrument of domestication within the broader technogender ecology. GenAI becomes a device for industrial-level fabrication of subjectivity, teaching subjects how to be legible to other (generative) AI systems. Such a statement raises political questions: Who becomes legible under the hallucinated subjects? And what possibilities remain for subjects who cannot (or refuse to) align with their hallucinated doubles?

Legibility and the Flattening of Queerness

Algorithmic publics demand computational legibility (Cheney-Lippold, 2011). The infrastructures that gather and process traces are not indifferent to form; they require identities that can be coherently tied back to individual subjects, aggregated into populations, and projected into the future. What these systems “want,” in a political-economic sense, is ever more data that can be organized into coherent patterns suitable for prediction and control: to sell things that will be bought, and to preempt forms of conduct that might disrupt existing orders—whether through consumer defection, protest, or other forms of collective change.

This demand for legibility is not new. Glissant (1997) famously defends a *right to opacity*: the right of subjects, particularly in colonized contexts, not to be fully known, translated, or reduced to categories legible to dominant epistemologies. Data feminism scholars such as D’Ignazio and Klein (2020) similarly emphasize that demands for transparency and legibility often mask power differentials, as legibility is frequently demanded of the marginalized, not of the powerful. Bey (2022) extends this politics of opacity through the lens of Black trans feminism, arguing that Blackness and transness name “fugitive postures,” modes of being that exceed and refuse the categorical demands of dominant technological systems. In Bey’s (2022) account, the *cistem* (a portmanteau of cisgender and system)—the interlocking apparatus of cisnormativity and systemic power—fails precisely because it cannot accommodate the radical indeterminacy of Black trans life. This fugitivity is not a lack but a generative refusal of the extractive legibility that computational systems demand. What is distinctive in the GenAI moment is that legibility is increasingly defined in computational, and increasingly narrative, terms. It is not enough to be legible to other humans; one must be narratable to GenAI models. Black feminist theorists have extended this politics of opacity into frameworks of refusal and fugitivity. Bey (2022) argues that the *cistem* can only be disrupted through practices that refuse its terms of legibility entirely. Similarly, Campt (2017) theorizes Black feminist futurity as a practice of refusal that operates not through loud protest but through quiet, everyday acts of living otherwise. These frameworks offer a more politically grounded account of what is at stake in opacity than resistance to transparency alone: They name a mode of being that exceeds and undermines the coherence infrastructure from within.

But the *algorithmic public* does not distribute this possibility equally. Its latent diffusion models treat Whiteness and cisnormativity as prototypical, such that queerness and racial difference appear as (sometimes hypervisible) amendments to a default rather than independent categories (Birhane, 2021). Alignment processes push models to produce consistent, noncontradictory representations of persons; user interfaces encourage people to present themselves in ways that are predictable and easily classifiable. Within this coherence infrastructure, incoherence is treated as an error to be corrected or as noise to be smoothed out to fit the dimensionality of latent space—not as a legitimate way of being.

For queer and trans subjects, this is a familiar dynamic. Multiplicity, ambiguity, transition, and opacity are often experienced as central to life and survival. Yet in data infrastructures, these dimensions appear as “uncertain ground truth,” “label noise,” or “non-actionable” complexity. Bias discourse in AI frequently assumes that there is a correct label for each person (Keyes, 2018), which can be recovered and aligned with through better annotation and modeling. From this perspective, harm is understood narrowly as misclassification. We argue that this framing misses a deeper issue: what we might call *coercive ontology*. Systems project normative assumptions about binary gender, stability, and coherence onto subjects, and treat deviations as problems to be resolved. Queer and trans ontologies often refuse the premise that identity should be extractable and stable. Even “fixing” models to better align with additional categories, such as by adding more options, improving pronoun recognition, or refining labels, can be ontologically violent when those categories themselves are instruments of control. That is, queer inclusion can function as *capture*. As queer and non-normative identities become more visible in training data, they open new frontiers for monetization and governance. Latent spaces can be “queerified” in representational terms, through more categories, gradients, and inclusive avatars, while still enforcing extractive and colonial logics that treat identity as a resource to be profiled, predicted, and exploited.

Who Is Afraid of Gender?

The coherence infrastructure described previously shares its governing conditions with a broader reactionary political moment. Butler (2024) argues that anti-gender movements construct a *phantasm*, a fictional, threatening “gender ideology,” to defend natural categories of sex against evidence that they are neither fixed nor natural. We see GenAI coherence infrastructures as doing the same work computationally. The rise of feminized AI-assistants and the idea of a “wife drought” are not incidental. The corporate executives driving GenAI development have been among the most visible advocates of pronatalist politics (such as Elon Musk’s public preoccupation with birth rates, Marc Andreessen’s techno-optimist manifesto, and the broader natalist turn of the Silicon Valley investor class), and the systems they build reflect this. GenAI that performs domestic and affective labor does not challenge the arrangement in which women’s reproductive and caring labor underwrites men’s productivity; rather, it extends that arrangement into the computational infrastructure, making it cheaper, scalable, and harder to refuse.

The same pronatalist logic underwrites the pressure to fix categories of sex and gender in law and policy (Haimson, 2025). Binary-enforcing biometrics and anti-trans classification compress all gender multiplicity into resolvable data, but the targeting of trans women specifically reveals what is being defended: a category of womanhood anchored in reproductive biology. The reproductive subject becomes the infrastructural ground zero of technogender, the point at which biological data is most aggressively captured and most consequentially deployed. Those who deviate from dominant norms of masculinity bear some of the greatest weight of the coherence demands of generative systems. The feminized, the queer, and the trans*: These are some of the subjects whose coherence has been decided in advance, by the same political conditions simultaneously being fought out in law, policy, and the molecular arrangements of everyday life.

Ungovernability, Refusal, and Relationality

If the hallucinated machinic subject becomes folded into self-narration, decision-making, and increasingly governance and surveillance, then the hallucinated human subject is constructed by the system regardless of individual consent to its terms. For those whose lives exceed or refuse these rigid categories, what is returned is not reflection but correction. Hallucination becomes a coercive projection: Models stabilize identities where they are shifting, refused, or only partially articulated; they resolve gender and sexuality even when people cannot or do not wish to.

We are thus left with a tension of algorithmic governance that specifically targets queer multiplicity. Queer self-formation involves performative negotiations of norms (Butler, 1993) that resist encoding in (training) data, since the extraction and narrative synthesis of intimate data inherently flattens or erases the contradictions of lived experience into latent approximations. Linked infrastructures (increasingly governing everything from visa approvals to healthcare access) seek to bind these internal statistical fictions back to flesh-and-bone bodies. In this way, GenAI's insistence on coherent identity is structurally at odds with queer existence.

Further, if the hallucinated subject is a technological fabrication, the question is not how to reclaim or see reflected an "authentic" self, but how to interrupt this construction. Preciado (2008/2013, p. 353) suggests that we must accept the radically techno-constructed and plastic nature of identity by adopting what he calls the "principle of the auto-guinea pig": a positioned, responsible corporal political practice where one becomes "the lab rat in her or his own laboratory." In this context, resistance is not found in an "outside," but in the experimental manipulation of the very semiotكنولوجies that seek to govern us. This auto-experimental refusal recuperates the spirit of early anonymous online spaces where, as Turkle (1995) observed, virtual environments functioned as laboratories for identity exploration. However, the GenAI user today plays inside a much more rigid coherence regime; the system is simultaneously the stage and the censor. Despite this, trans* users have been especially resourceful in finding escape lines. This ranges from the speculative AI image-making documented by Kafer (2025), where marginalized communities "fabricate" inclusive visual histories to fill the gaps of the archive, to Haimson's (2025) study of community-built technologies designed to route around hostile platforms. These practices represent an ontological refusal: a challenge to the premise that identity must be extractable, coherent, and computable. These practices are not merely defensive responses to misclassification; they constitute forms of self-determination through which trans subjects expose the contingency of computational gender itself, demonstrating that the categories enforced by these systems are neither natural nor necessary but technically and politically constructed.

From this angle, the hallucinated subjects are not a technical aberration but an extreme form of industrial-level technical fabrication of predictable subjectivity. The question, then, is not whether subjects are technically constructed, but how subjects might sabotage the coherence infrastructures that seek to preempt their futures. Queer resistance, in this context, is an ontological refusal of governable subjectivity. Possibilities include strategic opacity, trace manipulation, and platform refusal. More radical interventions might involve adversarial inputs that destabilize latent categories, collective data practices that refuse individual profiling, or counter-infrastructures that do not require identity resolution. The goal is not a better mirror but a fundamentally different relationship to the apparatus of identity production.

Critically, ungovernability does not arise from personal opt-outs alone. It instead lies in collective practices that exceed the imagination of latent space and ID-linked governance: new forms of solidarity, community-built datasets and models, spaces where opacity is a feature. Such "alternative relationalities" grounded in care, interdependence, and shared precarity (Iapaolo & Lynch 2026) are extensions of what marginalized communities have always done: build infrastructures of survival that refuse the terms imposed by oppressive systems. To cultivate ungovernable subjectivities under increasingly technofascist conditions is to inhabit and twist GenAI infrastructures without being fully determined by them: to become "both the terminal of one of the apparatuses of neoliberal governmentality and the vanishing point through which the system's power to

control escapes” (Preciado, 2008/2013, p. 155). Queer and trans refusal thus names subjectivities who do not just absorb GenAI systems’ demands but glitch them—exceeding predictability and opening trajectories of life that governance cannot fully script.

Conclusion

Hallucination, in technical literature, names a model failure. It is a deviation from a grounded truth: A GenAI model made a claim, and what it claimed was false. This definition keeps hallucination at the level of propositional accuracy. Yet the same generative process also operates at the level of subject formation. The problem, then, is not that AI systems occasionally produce incorrect content about people; it is that they continuously produce people through the hallucination of subjects: unstable figures of personhood, gendered in their fabrication, assembled from exploitative and unauthorized data labor, and optimized toward a coherence aligned with dominant norms, regardless of whether that coherence bears any relation to the life being modeled.

Throughout this article we have tracked that subjectivity production at two poles: (a) the hallucinated human subject, the modeled “you,” inferred from latent proximities and stabilized into a legible identity that the GenAI system can address, predict, and govern; and (b) the hallucinated machinic subject, the performed “I,” a chatbot persona optimized to sustain narrative continuity and affective authority, producing the impression of a “self” precisely where there is none. Between them, there is a feedback loop that is also a coherence regime. Each pole reinforces the other, and both work to domesticate the lived, fractured, relational self of personhood that entered the interaction in the first place. The question we close with is not whether AI systems can be made more accurate. It is what kinds of subjectivities these systems make it possible (or impossible) to inhabit.

Future work, we suggest, must grapple with GenAI as a coherence machine that hallucinates both human and machinic subjects—flattening lived, fractured selves into legible, technogendered identities optimized for prediction. Methodologically, this requires ways of making generative algorithmic publics visible and contestable: tracing how traces are extracted and circulate, how latent subjects are assembled, how machinic personas are tuned, and how these processes differentially affect those at the sharpest edges of technogendered control—the corpus pornographicus, bodies deprived of civic condition. Politically and imaginatively, it calls for designing and defending spaces where opacity, multiplicity, and refusal are not pathologized but recognized as conditions of livable lives (Butler, 2004).

The question “who is the hallucinated subject?” does not have a single answer. It points at once to users’ and models’ datafied fictions, imaginary subjectivities that are predictable, both preceding and outliving them. Our aim has not been to choose between these figures, but to show how, under GenAI, the boundary between them is continually crossed in the fabrication of hallucinatory selves: how technologies increasingly get under our skin, how our bodies increasingly become infrastructure.

The authors used an AI-assisted language editing tool to support manuscript preparation. This use was undertaken in accordance with the authors’ disclosure obligations and reflects a specific accessibility need: One of the authors has dyslexia, and AI-assisted language editing was used to support clarity of written expression, proofreading, and consistency of spelling and phrasing. The authors retain full responsibility for the intellectual content, arguments, analysis, and conclusions of this article, all of which are entirely their own. No AI tool was used to generate substantive content, develop arguments, conduct research, or draft original text.

References

- Althusser, L (1971). Ideology and ideological state apparatuses. In A Scott & K Nash (Eds.), *New critical writings in political sociology* (pp. 299–340). Routledge. <https://doi.org/10.4324/9781003572923-18>
[Google Scholar](#) [Google Preview](#) [WorldCat](#) [COPAC](#)
- Amoore, L, Campolo, A, Jacobsen, B, & Rella, L (2024). A world model: On the political logics of generative AI. *Political Geography*, 113, Article 103134. <https://doi.org/10.1016/j.polgeo.2024.103134>
[Google Scholar](#) [WorldCat](#)
- Anthropic. (2026a). *Claude's constitution*. <https://www.anthropic.com/constitution>.
[Google Scholar](#) [Google Preview](#) [WorldCat](#) [COPAC](#)
- Anthropic. (2026b). *Model system cards*. <https://www.anthropic.com/system-cards>.
[Google Scholar](#) [Google Preview](#) [WorldCat](#) [COPAC](#)
- Asia, M G (2025). The quiet cost of emotional labor. In M Miceli, A Dinika, K Kauffman, C S Wagner, & L Sachenbacher (Eds.), *The quiet cost of emotional labor*. <https://data-workers.org/michael>.
[Google Scholar](#) [Google Preview](#) [WorldCat](#) [COPAC](#)
- Atanasoski, N, & Vora, K (2019). *Surrogate humanity: Race, robots, and the politics of technological futures*. Duke University Press.
[Google Scholar](#) [Google Preview](#) [WorldCat](#) [COPAC](#)
- Bailey, M (2021). *Misogynoir transformed: Black women's digital resistance*. NYU Press.
[Google Scholar](#) [Google Preview](#) [WorldCat](#) [COPAC](#)
- Barad, K (2003). Posthumanist performativity: Toward an understanding of how matter comes to matter. *Signs: Journal of Women in Culture and Society*, 28(3), 801–831. <https://doi.org/10.1086/345321>
[Google Scholar](#) [WorldCat](#)
- Barkallah, M, & Zytco, D (2026, April). “I wanted them to think that I wrote that”: AI-generated self-presentation on dating apps and implications of non-disclosure on informed consent. In *Proceedings of the 2026 CHI Conference on Human Factors in Computing Systems (CHI'26)* (pp. 1–18). Association for Computing Machinery, New York, NY, USA, Article 112.
<https://doi.org/10.1145/3772318.3791593>
[Google Scholar](#) [Google Preview](#) [WorldCat](#) [COPAC](#)
- Bender, E M, Gebru, T, McMillan-Major, A, & Shmitchell, S (2021). On the dangers of stochastic parrots: Can language models be too big? In *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency* (pp. 610–623). Association for Computing Machinery. <https://doi.org/10.1145/3442188.3445922>
[Google Scholar](#) [Google Preview](#) [WorldCat](#) [COPAC](#)
- Bey, M (2022). *Cistem failure: Essays on Blackness and cisgender*. Duke University Press.
[Google Scholar](#) [Google Preview](#) [WorldCat](#) [COPAC](#)
- Berger, E (2025, November 28). After a teddy bear talked about kink, AI watchdogs are warning parents against smart toys. *The Guardian*. www.theguardian.com/technology/2025/nov/28/artificial-intelligence-smart-toys
[Google Scholar](#) [Google Preview](#) [WorldCat](#) [COPAC](#)
- Bhandari, A, & Bimo, S (2022). Why's everyone on TikTok now? The algorithmized self and the future of self-making on social media. *Social Media + Society*, 8(1), Article 20563051221086241. <https://doi.org/10.1177/20563051221086241>
[Google Scholar](#) [WorldCat](#)
- Birhane, A (2021). Algorithmic injustice: A relational ethics approach. *Patterns*, 2(2), Article 100205.
<https://doi.org/10.1016/j.patter.2021.100205>

[Google Scholar](#) [WorldCat](#)

Bolis, D, Birhane, A, Pollak, T, & Schilbach, L (2026). *In defense of misattunement: Hyperalignment in human–AI interactions* [Preprint]. PsyArXiv. https://doi.org/10.31234/osf.io/nqjke_v1

[Google Scholar](#) [Google Preview](#) [WorldCat](#) [COPAC](#)

Butler, J (1993). *Bodies that matter: On the discursive limits of “sex.”* Routledge.

[Google Scholar](#) [Google Preview](#) [WorldCat](#) [COPAC](#)

Butler, J (2004). *Undoing gender.* Routledge.

[Google Scholar](#) [Google Preview](#) [WorldCat](#) [COPAC](#)

Butler, J (2006). *Gender trouble: Feminism and the subversion of identity.* Routledge.

[Google Scholar](#) [Google Preview](#) [WorldCat](#) [COPAC](#)

Butler, J (2024). *Who’s afraid of gender?* Farrar, Straus and Giroux.

[Google Scholar](#) [Google Preview](#) [WorldCat](#) [COPAC](#)

Campt, T M (2017). *Listening to images.* Duke University Press.

[Google Scholar](#) [Google Preview](#) [WorldCat](#) [COPAC](#)

Card, S K, Moran, T P, & Newell, A (2018). *The psychology of human-computer interaction* (S K Card, Ed.). CRC Press.

<https://doi.org/10.1201/9780203736166>

[Google Scholar](#) [Google Preview](#) [WorldCat](#) [COPAC](#)

Castells, M (2009). *The rise of the network society* (2nd ed). John Wiley & Sons.

[Google Scholar](#) [Google Preview](#) [WorldCat](#) [COPAC](#)

Cheney-Lippold, J (2011). A new algorithmic identity: Soft biopolitics and the modulation of control. *Theory, Culture & Society*, 28(6), 164–181. <https://doi.org/10.1177/0263276411424420>

[Google Scholar](#) [WorldCat](#)

Clowes, R W, Gärtner, K, & Theiner, G (2026). The mind-technology problem in the age of GenAI: Introduction to the special issue. *Social Epistemology*, 40(1), 1–15. <https://doi.org/10.1080/02691728.2025.2574296>

[Google Scholar](#) [WorldCat](#)

Crawford, K (2021). *Atlas of AI: Power, politics, and the planetary costs of artificial intelligence.* Yale University Press.

[Google Scholar](#) [Google Preview](#) [WorldCat](#) [COPAC](#)

Dash, A, Das, S, Kirsten, E, Wu, Q, Karnam, S K, Gummadi, K P, . . . Zannettou, S (2026, April). The algorithmic self-portrait: Deconstructing memory in ChatGPT. In *Proceedings of the ACM Web Conference 2026* (pp. 3471–3482). Association for Computing Machinery, New York, NY, USA. <https://doi.org/10.1145/3774904.3792671>

[Google Scholar](#) [Google Preview](#) [WorldCat](#) [COPAC](#)

De Lauretis, T (2010). *Technologies of gender: Essays on theory, film, and fiction.* Indiana University Press.

[Google Scholar](#) [Google Preview](#) [WorldCat](#) [COPAC](#)

Deleuze, G (1992). Postscript on the societies of control. *October*, 59, 3–7.

[Google Scholar](#) [WorldCat](#)

Dev, S, Monajatipoor, M, Ovalle, A, Subramonian, A, Phillips, J M, & Chang, K -W (2021). Harms of gender exclusivity and challenges in nonbinary representation in language technologies. In *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing* (pp. 1968–1994). Association for Computational Linguistics.

<https://doi.org/10.18653/v1/2021.emnlp-main.150>

[Google Scholar](#) [Google Preview](#) [WorldCat](#) [COPAC](#)

D'Ignazio, C, & Klein, L F (2020). *Data feminism*. MIT Press.

[Google Scholar](#) [Google Preview](#) [WorldCat](#) [COPAC](#)

Dourish, P (2004). *Where the action is: The foundations of embodied interaction*. MIT Press.

[Google Scholar](#) [Google Preview](#) [WorldCat](#) [COPAC](#)

Drage, E, & Frabetti, F (2024). The performativity of AI-powered event detection: How AI creates a racialized protest and why looking for bias is not a solution. *Science, Technology, & Human Values*, 49(5), 1045–1072.

<https://doi.org/10.1177/01622439231164660>

[Google Scholar](#) [WorldCat](#)

DuBrin, R, & Gorham, A E (2021). Algorithmic interpellation. *Constellations*, 28(2), 176–191. <https://doi.org/10.1111/1467-8675.12568>

[Google Scholar](#) [WorldCat](#)

Gillespie, T (2024). Generative AI and the politics of visibility. *Big Data & Society*, 11(2), Article 20539517241252131.

<https://doi.org/10.1177/20539517241252131>

[Google Scholar](#) [WorldCat](#)

Glissant, E (1997). *Poetics of relation* (B. Wing, Trans.). University of Michigan Press. <https://doi.org/10.3998/mpub.10257>

[Google Scholar](#) [Google Preview](#) [WorldCat](#) [COPAC](#)

Gray, M L, & Suri, S (2019). *Ghost work: How to stop Silicon Valley from building a new global underclass*. Houghton Mifflin Harcourt.

[Google Scholar](#) [Google Preview](#) [WorldCat](#) [COPAC](#)

Goffman, E (2007). *The presentation of self in everyday life (Reprint)*. Penguin Books.

[Google Scholar](#) [Google Preview](#) [WorldCat](#) [COPAC](#)

Gold, H (2025, September 5). They thought they were making technological breakthroughs. It was an AI-sparked delusion. *CNN*.

<https://edition.cnn.com/2025/09/05/tech/ai-sparked-delusion-chatgpt>. Accessed May 27, 2026.

[Google Scholar](#) [WorldCat](#)

Guattari, F (2013). *Schizoanalytic cartographies* (A. Goffey, Trans.). Bloomsbury.

[Google Scholar](#) [Google Preview](#) [WorldCat](#) [COPAC](#)

Haggerty, K D, & Ericson, R V (2000). The surveillant assemblage. *British Journal of Sociology*, 51(4), 605–622.

<https://doi.org/10.1080/00071310020015280>

[Google Scholar](#) [PubMed](#) [WorldCat](#)

Haimson, O (2025). *Trans technologies*. MIT Press.

[Google Scholar](#) [Google Preview](#) [WorldCat](#) [COPAC](#)

Han, J H (2024, April 1). *The advent of the bespoke AI era*. Samsung Global Newsroom.

<https://news.samsung.com/global/editorial-the-advent-of-the-bespoke-ai-era>. Accessed 27 May 2026.

[Google Scholar](#) [Google Preview](#) [WorldCat](#) [COPAC](#)

Haraway, D (1991). *Simians, cyborgs and women: The reinvention of nature*. Routledge.

<https://web.archive.org/web/20120214194015/http://www.stanford.edu/dept/HPS/Haraway/CyborgManifesto.html>

[Google Scholar](#) [Google Preview](#) [WorldCat](#) [COPAC](#)

Hilty, K [@kendrahilty]. (2025, March 3). *Part 1: I fell in love with my psychiatrist* [Video]. TikTok.

<https://www.tiktok.com/@kendrahilty/video/7533994100492700942>

[Google Scholar](#) [Google Preview](#) [WorldCat](#) [COPAC](#)

Hook, S, & Sassine, J (2026). He, her, they robot: Gendered AI chatbots and the proliferation of technologically facilitated violence. *Griffith Law Review*, 35(1), 1–20. <https://doi.org/10.1080/10383441.2026.2621579>
[Google Scholar](#) [WorldCat](#)

Huang, L, Yu, W, Ma, W, Zhong, W, Feng, Z, Wang, H, Chen, Q, Peng, W, Feng, X, Qin, B, & Liu, T (2025). A survey on hallucination in large language models: Principles, taxonomy, challenges, and open questions. *ACM Transactions on Information Systems*, 43(2), Article 42. <https://doi.org/10.1145/3703155>
[Google Scholar](#) [WorldCat](#)

Iapaolo, F (2026). The latent subject: AI, recognition, and the politics of latent space. *Political Geography*, 47, Article 103500. <https://doi.org/10.1016/j.polgeo.2026.103500>
[Google Scholar](#) [WorldCat](#)

Iapaolo, F, & Lynch, C R (2026). Rethinking urban AI: A critical posthumanist perspective on intelligence, autonomy, and agency. *Urban Geography*, 47(1), 65–87. <https://doi.org/10.1080/02723638.2025.2510812>
[Google Scholar](#) [WorldCat](#)

Jarrett, K (2016). *Feminism, labour and digital media: The digital housewife*. Routledge.
[Google Scholar](#) [Google Preview](#) [WorldCat](#) [COPAC](#)

Kafer, G (2025). Speculating in latent space: Visibility politics and the impasse of representation in generative AI. *Journal of Cinema and Media Studies*, 64(2), 102–125. <https://doi.org/10.1353/cj.2025.a951188>
[Google Scholar](#) [WorldCat](#)

Kalai, A T, Nachum, O, Vempala, S S, & Zhang, E (2025). *Why language models hallucinate* [Preprint]. arXiv. <https://doi.org/10.48550/arXiv.2509.04664>
[Google Scholar](#) [Google Preview](#) [WorldCat](#) [COPAC](#)

Keyes, O (2018). The misgendering machines: Trans/HCI implications of automatic gender recognition. *Proceedings of the ACM on Human-Computer Interaction*, 2(CSCW), 1–22. <https://doi.org/10.1145/3274357>
[Google Scholar](#) [Google Preview](#) [WorldCat](#) [COPAC](#)

Keyes, O (2023). Inconclusion. In M Klippahhn-Karge, A-K Koster, & S Morais Dos Santos Bruss (Eds.), *Queer reflections on AI* (pp. 179–186). Routledge. <https://doi.org/10.4324/9781003357957-15>
[Google Scholar](#) [Google Preview](#) [WorldCat](#) [COPAC](#)

Kingma, D P, & Welling, M (2014). *Auto-encoding variational Bayes* [Preprint]. arXiv. <https://arxiv.org/abs/1312.6114>
[Google Scholar](#) [Google Preview](#) [WorldCat](#) [COPAC](#)

Kitchin, R (2014). *The data revolution: Big data, open data, data infrastructures and their consequences*. Sage.
[Google Scholar](#) [Google Preview](#) [WorldCat](#) [COPAC](#)

Lin, Y -Z, Shih, B P -J, Chien, H -Y A, Satam, S, Pacheco, J H, Shao, S, Salehi, S, & Satam, P (2026). *Exploring Silicon-Based societies: An early study of the Moltbook agent community* [Preprint]. arXiv. <https://doi.org/10.48550/arXiv.2602.02613>
[Google Scholar](#) [Google Preview](#) [WorldCat](#) [COPAC](#)

MacKenzie, D A, & Wajcman, J (Eds.). (2011). *The social shaping of technology* (2nd ed.). Open University Press.
[Google Scholar](#) [Google Preview](#) [WorldCat](#) [COPAC](#)

Magee, L, Arora, V, & Munn, L (2023). Structured like a language model: Analysing AI as an automated subject. *Big Data & Society*, 10(2), Article 20539517231210273. <https://doi.org/10.1177/20539517231210273>
[Google Scholar](#) [WorldCat](#)

Maleki, N, Padmanabhan, B, & Dutta, K (2024). AI hallucinations: A misnomer worth clarifying. In *2024 IEEE Conference on Artificial Intelligence (CAI)* (pp. 133–138). IEEE. <https://doi.org/10.1109/CAI59869.2024.00033>

[Google Scholar](#) [Google Preview](#) [WorldCat](#) [COPAC](#)

Mead, G H, Morris, C W, Huebner, D R, & Joas, H (2015). *Mind, self, and society* (Definitive edition). University of Chicago Press.

[Google Scholar](#) [Google Preview](#) [WorldCat](#) [COPAC](#)

Mohamed, S, Png, M -T, & Isaac, W (2020). Decolonial AI: Decolonial theory as sociotechnical foresight in artificial intelligence. *Philosophy & Technology*, 33(4), 659–684. <https://doi.org/10.1007/s13347-020-00405-8>

[Google Scholar](#) [WorldCat](#)

Neff, G, & Nagy, P (2025). The quasi-domestication of social chatbots: The case of Replika. *New Media & Society*, 27(10), 5508–5524. <https://doi.org/10.1177/14614448251359218>

[Google Scholar](#) [WorldCat](#)

OpenAI. (2025). *Sycophancy in GPT-4o: What happened and what we're doing about it*. <https://openai.com/index/sycophancy-in-gpt-4o/>.

[Google Scholar](#) [Google Preview](#) [WorldCat](#) [COPAC](#)

Østergaard, S D (2023). Will generative artificial intelligence chatbots generate delusions in individuals prone to psychosis? *Schizophrenia Bulletin*, 49(6), 1418–1419. <https://doi.org/10.1093/schbul/sbad128>

[Google Scholar](#) [PubMed](#) [WorldCat](#)

Østergaard, S D (2025). Generative artificial intelligence chatbots and delusions: From guesswork to emerging cases. *Acta Psychiatrica Scandinavica*, 152(4), 257–259. <https://doi.org/10.1111/acps.70022>

[Google Scholar](#) [PubMed](#) [WorldCat](#)

Ovalle, A, et al. (2023). Factoring the matrix of domination: A critical review and reimagination of intersectionality in ai fairness. In *Proceedings of the 2023 AAAI/ACM Conference on AI, Ethics, and Society* (pp. 496–511). Association for Computing Machinery. <https://doi.org/10.1145/3600211.3604705>

[Google Scholar](#) [Google Preview](#) [WorldCat](#) [COPAC](#)

Preciado, P B (2013). *Testo junkie: Sex, drugs, and biopolitics in the pharmacopornographic era* (B. Benderson, Trans.). Feminist Press at The City University of New York. (Original work published 2008)

[Google Scholar](#) [Google Preview](#) [WorldCat](#) [COPAC](#)

Saslow, E (2026, February 12). To stay in her home, she let in an A.I. robot. *New York Times*.

<https://www.nytimes.com/2026/02/12/us/elliq-ai-robot-senior-companion.html>.

[Google Scholar](#) [Google Preview](#) [WorldCat](#) [COPAC](#)

Scheuerman, M K, Wade, K, Lustig, C, & Brubaker, J R (2020). How we've taught algorithms to see identity: Constructing race and gender in image databases for facial analysis. *Proceedings of the ACM on Human-Computer Interaction*, 4(CSCW1), Article 58.

<https://doi.org/10.1145/3392866>

[Google Scholar](#) [WorldCat](#)

Strengers, Y, & Kennedy, J (2020). *The smart wife: Why Siri, Alexa, and other smart home devices need a feminist reboot*. MIT Press.

[Google Scholar](#) [Google Preview](#) [WorldCat](#) [COPAC](#)

Suchman, L (2007). *Human-machine reconfigurations: Plans and situated actions* (2nd ed.). Cambridge University Press.

[Google Scholar](#) [Google Preview](#) [WorldCat](#) [COPAC](#)

Sui, P, Duode, E, Wu, S, & So, R (2024). Confabulation: The surprising value of large language model hallucinations. In L -W Ku, A Martins, & V Srikumar (Eds.), *Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)* (pp. 14274–14284). Association for Computational Linguistics. <https://doi.org/10.18653/v1/2024.acl-long.770>

[Google Scholar](#) [Google Preview](#) [WorldCat](#) [COPAC](#)

Tonmoy, S M T I, Zaman, S M M, Jain, V, Rani, A, Rawte, V, Chadha, A, & Das, A (2024). *A comprehensive survey of hallucination mitigation techniques in large language models* [Preprint]. arXiv. <https://doi.org/10.48550/arXiv.2401.01313>

[Google Scholar](#) [Google Preview](#) [WorldCat](#) [COPAC](#)

Turkle, S (1995). *Life on the screen: Identity in the age of the Internet*. Simon and Schuster.

[Google Scholar](#) [Google Preview](#) [WorldCat](#) [COPAC](#)

Verhagen, M D, Stroebel, B, Liu, T, Liu, L T, & Salganik, M J (2025). *The Book of Life approach: Enabling richness and scale for life course research* [Preprint]. arXiv. <https://doi.org/10.48550/arXiv.2507.03027>

[Google Scholar](#) [Google Preview](#) [WorldCat](#) [COPAC](#)

Walter, Y (2025). Artificial influencers and the dead internet theory. *AI & SOCIETY*, 40(1), 239–240. <https://doi.org/10.1007/s00146-023-01857-0>

[Google Scholar](#) [WorldCat](#)

Wing, J M (2006). Computational thinking. *Communications of the ACM*, 49(3), 33–35. <https://doi.org/10.1145/1118178.1118215>

[Google Scholar](#) [WorldCat](#)

Wong, J, & Kim, J (2023). *ChatGPT is more likely to be perceived as male than female* [Preprint]. arXiv.

<https://doi.org/10.48550/arXiv.2305.12564>

[Google Scholar](#) [Google Preview](#) [WorldCat](#) [COPAC](#)

Zhi-Xuan, T, Carroll, M, Franklin, M, & Ashton, H (2025). Beyond preferences in AI alignment. *Philosophical Studies*, 182(7), 1813–1863. <https://doi.org/10.1007/s11098-024-02249-w>

[Google Scholar](#) [WorldCat](#)

Zuboff, S (2019). *The age of surveillance capitalism: The fight for a human future at the new frontier of power*. Profile books.

[Google Scholar](#) [Google Preview](#) [WorldCat](#) [COPAC](#)